



Autonomous reciprocity: context matters

Karamjit S. Gill¹

Accepted: 21 February 2022 / Published online: 28 February 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

In continuing the debate on actionable ethics, the special issue on Critical Robotics (this volume) explores issues of ethics, sustainability, and responsibility in social robotics. Within the context of social robots and the ethics of care, we note possible effects of nudging in reciprocal relationships between humans and robots. Further, we are alerted to the danger of designing social robots for reciprocity where reciprocity may be used as an instrumental value to enhance acceptability of the robot, and this is ethically questionable. However, in contrast, we also learn how humans develop empathic responses to robots. This argument on the ethical reciprocity draws upon the philosophy of the Danish theologian K. E. Løgstrup, that human empathy is inherently good because it turns people away from their own self-focus (inturnedness), and this concept of empathy applies also to relations with robots. Although it is acknowledged that reciprocity is indeed a component of moral development, and is in no way harmful in itself. It is, however, uncertain whether reciprocity fostered in Human Robot Interaction (HRI) would transfer to human–human interaction where it would provide the most benefit. It is thus much better to focus on fostering reciprocity among humans to facilitate human–robot interactions. We are asked to pay attention to the debate on unintended or undesirable consequences of empathic responses of human to robots, for example the potential for malicious intent and exploitation in robot design and development in the name of ethical and socio-emotional relationships with robots. The notions of human–robot reciprocity and empathetic interaction highlight the oversimplification of social care and service practices in the design of human likeness in robots as social companions. The core premise of this articulation of reciprocity is that sociality is not something that can be a property of a machine, but is rather something that is enacted in an encounter, or an evolving relationship, between a human and a machine. If

this is the case, then we should focus on the enactment of empathic social agency, rather than its representation, in the design of social robots.

We note how the representation of empathetic reciprocity is propagated in the design of social robotics for the care and service sectors, for example in the therapy and care of dementia patients, robot companions for older adults living at home. The idea of robots providing services that we would otherwise expect from humans forces us to think about the aspects of these services that may, and may not, be replaceable. Here the technologies that promise remedies to human vulnerabilities seem very enticing, and this faith in technological solution of social problems leads to an oversimplification of the role of humans in care and service work, or a reduction in the complexity of the tasks that they carry out. By depicting older adults as dependent, fragile and vulnerable people, renders them as ‘potentially burdensome care recipients’, and robot technologies are presented as an optimal solution to this social problem. In a similar vein, the roles of caregivers and care-receivers, and care practices are deconstructed into tasks to fit well-defined technical problems. This leads to an incremental mechanization of care, rather than to a more holistic understanding of it. This oversimplification of social care, rooted in misconceptions about the provision of care, the process of ageing, affective labor in professional service work, can influence the design and implementation of social robots. Although there is a deep concern about the potential replacement of human care providers with robotic technologies, the introduction of robots as complementary technologies in social settings raises important ethical questions.

The special issue on ‘Drones, robots and perceived autonomy’ (this volume) extends and draws our attention to the debate on ethics to ethics and autonomy, where human subjects attribute *autonomy* to their experience of artificial devices. The idea of autonomy, rooted in Western philosophy of Aristotle and Kant, widely assumes that any perceivable action has a ‘source’ that centers on an actor/agent. In some undefined sense, humans (and all living beings) are taken to act autonomously, and this view thus has consequences for

✉ Karamjit S. Gill
editoraisoc@yahoo.co.uk

¹ University of Brighton, Brighton, UK

living human beings. This perception of autonomy not only impinges on organizational, social and individual experiences and actions, but also on how we conceptualize AI devices such as predator drones as killer robots, and our roles as actors (and entities) and its implications for designers of such machines. The perceived autonomy is thus related to not only how autonomy is perceived but also how working with human–machine aggregates form a broader perspective of interactional and situational outcomes, socio-cognitive organization, culture and, thus, of the ethical issues that are central to AI. The question arises whether we could—indeed should build machines as moral actors, and in what ways those working in machine ethics treat the autonomy of artificial agents as quite unlike that of natural agents. If Kantian view of ethics and agency depends on the seat of reason or the *mind*, artificial moral agents (AMA) should not only be rational but also fundamentally subjective. From this perspective, ‘Kantian AMA’ would therefore pursue, not common interests or those of communities, but outcomes that are consistent with universal, individual and voluntarist reasoning. However, if we take Aristotelian tradition of ethics and agency in the sense that living human beings act ethically within a social context, then autonomy is not seen as intrinsic but, rather, fundamentally relational. In this case moral judgements can only be traced to the embodied socialization of a citizen. Depending upon whether we take Kantian view or Aristotelian view of autonomy, the AMAs would differ in evaluating what is good and appealing, on the one hand, to society as a whole and, on the other, to a rational grasp of what is right. This rests on the view that humans, at least, exhibit the autonomy of social beings, and further depends upon how we see AMAs, how we see their societal role and, how we regulate and motivate designers. The idea of the alignment of human values to the machine or AMAs continues to find its resonance of alignment of human skill and knowledge, in the special issue on Team autonomy and digital transformation (this volume), with a view to regulate and manage the team work in cooperation with internal and external partners and systems. This alignment creates a layer of dependency between the team and the multi-team environments work, thereby challenging autonomy of the team. Such an alignment is rooted in the techno-centric view that processes of multi-team environments can be converted into linear, sequential and reasonably predetermined routine forms. However, multi-team environments involve a much higher degree of ambiguity and nonlinearity in the conversion process. Additionally, all teams must operate in the

context of an environment. The type of environment for an autonomous team will vary a lot, and it may change over time. The environment will have a number of qualities and offer several types of relationships. Many of these relationships affect a team’s autonomy by producing dependencies between team and the environment or disruptions in relation to the team’s internal processes. The immediate environment is typically a host organization, but this may come in many variations. The team is structurally coupled to its host organization through a setup of instructions, tasks, regulations, communication structures and other factors, but dependencies on and influences from the environment can go far beyond this. They can be about normative expectations of various kinds, epistemological assumptions or mechanisms of power, and the source of such expectations is not limited to the host organization. The more variety and influence from the environment the team is exposed to, the more demanding it will be to maintain its autonomy.

It is proposed that one way to sustain team autonomy is to create a buffer (measure to protect team working against turbulence from the outside) between the team and the external environment. A key focus here is how teams, with the help of organizational processes, technology, new organizational forms, and time and space are able to create a buffer needed for maintaining team autonomy. Even this buffering faces new challenges when teams need to reach agreement with external experts, managers, stakeholders and other teams, when agile methods are used in large-scale scaling contexts. Further, quality concerns and the need for frequent releases require agile teams in large-scale projects to be aligned with the rest of the teams and the organization. One possible solution to maintain the buffer is to create a technical interface between teams and letting teams have control over their own code and services. This means if a team needs to access a service from another team, the service is accessed through a defined interface. When a team changes its software code, the change does not have any effect on other teams as long as the interface is the same. This strategy is often named API-centric architecture. The term API is an acronym for “Application Programming Interface”.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.